

# Macchine senza guidatore: considerazioni morali

Maurizio Balistreri

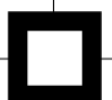
## Sommario

*I veicoli autonomi sono presentati come il futuro del mondo dei trasporti. Lo sviluppo di macchine intelligenti e senza guidatore pone importanti questioni morali. Queste macchine, ad esempio, potranno avere un giorno rilevanza morale ed avere anche diritti? Inoltre, una macchina capace di risolvere dilemmi può essere considerata un 'agente' morale? Infine, queste macchine dovranno essere programmate ad agire moralmente: ma a quali principi morali ed a quali valori vogliamo che esse facciano riferimento?*

## Abstract

*Autonomous vehicles are presented as the future of the world of transport. Their greatest advantage would be the reduction in deaths and injuries on the road due to accidents. The development of driverless vehicles raises important moral questions. Firstly, we are going to ask which duties we have towards intelligent machines and whether an autonomous driverless car may be considered a moral 'agent'. Further, the question must be faced of how to programme a machine to act morally and of which morality to take as a planning model to allow the autonomous car to respond to our expectations.*

**Keywords:** self-driving car, trolley problem, ethics, autonomous robots, algorithms, Artificial Intelligence



## 1. Introduzione: perché sviluppare macchine senza guidatore?

I veicoli autonomi vengono presentati come il futuro del mondo dei trasporti [1,2,3]. Quando avremo le prime macchine senza guidatore? Alcuni prevedono prima del 2020 [3,4,5,6], altri invece ritengono che ci vorranno non meno di altri due o tre decenni [7,8]. Si pensa poi che ai veicoli autonomi si arriverà passando per una fase intermedia, in cui i compiti del guidatore saranno piano piano delegati interamente ai robot intelligenti. Via via che le macchine diventeranno sempre più autonome, probabilmente il guidatore perderà progressivamente la dimestichezza con la guida e la capacità di affrontare situazioni impreviste, in quanto avrà sempre meno opportunità di esercitarsi [9]. Inoltre, i veicoli intelligenti ma soltanto parzialmente autonomi potrebbero risultare poco graditi o soddisfacenti per il consumatore, in quanto lo obbligano, negli viaggi in macchina, ad essere sempre attento alla strada. Per questa ragione, sarebbe meglio affidare la guida completamente al robot: ma almeno per il momento e per il prossimo futuro questa possibilità è da escludere, in quanto le macchine non hanno ancora la stessa prontezza di un essere umano nel percepire eventuali ostacoli, evitarli ed affrontare con dimestichezza nuove possibili circostanze. Per diventare competitive ed affermarsi, comunque, le macchine autonome non dovranno essere in grado di evitare qualsiasi danno a cose e/o persone [9]: basta che esse assicurino un numero di incidenti e di morti più basso di quello attuale. «Se i veicoli senza guidatore producessero meno incidenti sulla strada dei guidatori umani – ricorda l'OMS – questo darebbe una ragione forte di politica pubblica per muovere il prima possibile ad un traffico su strada interamente lasciato alle macchine autonome. Il costo umano del pedaggio stradale è tremendo. Il costo sociale ed economico del prendersi cura a breve e lungo termine delle vittime di incidenti stradali produce un consumo significativo delle risorse pubbliche» [10].

Con lo sviluppo di macchine autonome sempre più sicure la guida di una macchina potrebbe essere vietata oppure assicurare una macchina che può essere guidata potrebbe diventare così costoso che quasi nessuno potrebbe permettersi di acquistarla. In altri termini, come afferma Robert Sparrow «Se i veicoli senza conducenti non diventeranno sicuri, non sarà morale venderli. Una volta che saranno più sicuri dei conducenti umani, per quanto riguarda i rischi a terzi, allora dovrebbe essere illegale guidarle: a quel punto eventuali guidatori sarebbero simili a robot ubriachi» [9]. Per altro, nel caso in cui le macchine con guidatore causassero incidenti, anche le case di produzione potrebbero essere chiamate a risponderne, in quanto producono e mettono in vendita cose che possono arrecare un danno alle persone: «una parte offesa o i suoi familiari, o forse lo stesso governo potrebbe denunciare per la produzione di un prodotto non sicuro la società che ha venduto veicoli senza capacità di guida autonoma e/o quelle che permettono l'opzione della guida umano» [9].

Le macchine completamente autonome possono essere utili a qualsiasi persona: non soltanto ci farebbero risparmiare tantissimo tempo negli spostamenti, ma ci permetterebbero di impiegare il tempo del viaggio in attività

più 'produttive' della guida. Inoltre, esse ci permetterebbero di muoverci in condizioni di maggiore sicurezza. Mentre siamo al volante può capitare facilmente di distrarsi: ad esempio, possiamo distogliere lo sguardo dalla strada per rispondere ad una telefonata, per scrivere un messaggio, per cambiare la stazione della radio, per riallacciare la cintura di sicurezza, per osservare una pubblicità o buttare un occhio su un articolo di giornale. Sono piccole distrazioni che però sono pericolose sia per noi che per le altre persone. Le macchine autonome non avrebbero questi problemi, perché sarebbero programmate a monitorare la situazione del traffico e la strada per tutta la durata del viaggio: quello che accade all'interno della macchina o lungo la strada e che non riguarda la sicurezza non potrebbe catturare la loro attenzione – nemmeno per un secondo. Inoltre, le macchine autonome potrebbero risultare particolarmente vantaggiose a quelle persone affette da gravi disabilità o altri problemi e che oggi non guidano. Una persona non vedente, ad esempio, oggi può spostarsi per la città in macchina solamente se chiama un taxi o viene accompagnata da un adulto con la patente. E lo stesso vale per le persone anziane che non si muovono con facilità o con prontezza: anche loro non potrebbero mai mettersi al volante di una macchina o se lo facessero potrebbero mettere a rischio non soltanto la loro vita ma anche quella degli altri. Le macchine autonome non permetterebbero a queste persone di guidare, ma darebbero loro la stessa libertà di spostarsi per la città che oggi hanno le persone normodotate. È evidente, poi, che ci sarebbero vantaggi anche per le persone più giovani: esse non avrebbero più bisogno di chiedere un passaggio a mamma o papà, in quanto potrebbero farsi accompagnare a scuola o a casa dei loro amici dalla loro macchina (questa poi potrebbe ritornare a casa da sola per essere a disposizione degli altri membri della famiglia oppure andare a prendere le altre persone che l'hanno prenotata).

Le città cambierebbero perché non avremmo più bisogno di parcheggiare 'sotto casa': i veicoli potrebbero essere lasciati in qualsiasi luogo perché essi non avrebbero bisogno della nostra assistenza per trovare un'area di sosta o un parcheggio a lunga sosta. I parcheggi, inoltre, non dovrebbero trovarsi necessariamente nelle aree residenziali, in quanto quando ci serve una macchina potremmo probabilmente prenotarla. Questo significa che con il tempo potrebbero liberarsi importanti spazi abitativi: «Le torri per uffici, i cinema e i centri commerciali potrebbero eliminare i parcheggi che attualmente li circondano e li isolano dalla comunità circostante e potrebbero dotarsi invece di una piccola piazza o parcheggio dove le persone potrebbero essere lasciate e passate a prendere dai loro veicoli. Tutti questi vantaggi si realizzerebbero in misura ancora maggiore se, come abbiamo suggerito, è probabile che l'avvento di veicoli senza conducente ridurrà il numero totale di veicoli sulle strade» [9]. A livello stradale, inoltre, ci potrebbe essere un impatto positivo sul traffico, in quanto ci sarebbe un maggiore coordinamento tra le diverse macchine sia nei movimenti che per quanto riguarda la scelta del migliore percorso per raggiungere una destinazione: i veicoli, poi, potrebbero ridurre la distanza di sicurezza tra una macchina e l'altra, perché tutte le macchine saranno in grado di frenare o accelerare contemporaneamente [4,12]. Inoltre, per ridurre ulteriormente il traffico e favorire gli spostamenti, si potrebbe promuovere una

maggiore integrazione tra mezzi di trasporto diversi: «Combinati in questo modo, - afferma Luis Ferreras - i treni e i veicoli autonomi potrebbero offrire un servizio di trasporto pubblico 'porta a porta', superando il problema dell'ultimo miglio che attualmente tormenta i sistemi di trasporto pubblico» [13]. Infine, le strade sarebbero più belle non soltanto perché meno congestionate, ma anche perché non ci sarebbe più bisogno di segnali stradali, semafori e altre indicazioni stradali che molto spesso impediscono di guardare al paesaggio circostante. Quando necessario, le auto regoleranno autonomamente le loro velocità [14,15].

Lo sviluppo delle macchine senza guidatore ci interessa perché solleva importanti questioni morali che saranno al centro della nostra analisi nelle prossime pagine. Inizieremo a domandarci se automobili tecnologicamente avanzate, in grado non soltanto di riconoscere e risolvere eventuali dilemmi ma anche di interagire con l'utente umano possono avere soltanto un valore strumentale oppure anche valore morale. Se, in altri termini, saranno soltanto delle cose o avranno anche una loro dignità, per cui non potremo trattarle come vogliamo ma dovremo riconoscere loro diritti. Ci chiederemo, poi, se una macchina intelligente va considerata un agente morale, cioè un soggetto che possiamo considerare responsabile per quello che fa, perché capace non soltanto di distinguere ma anche di scegliere tra diversi corsi di azione: la risposta sembra semplice, ma come facciamo a giustificare le nostre conclusioni? È vero che una macchina intelligente (automobile o qualsiasi robot) agisce sulla base della programmazione che ha ricevuto: siamo sicuri però che noi possiamo controllare e scegliere le nostre azioni? Infine, c'è la questione che riguarda la morale a cui la macchina deve fare riferimento: lungo la strada che dovrà percorrere per trasportare i passeggeri a destinazione, l'automobile autonoma potrebbe trovarsi di fronte a importanti dilemmi morali. È importante che agisca bene, ma spetta a noi decidere che cosa vogliamo che faccia.

## **2. Una macchina senza guidatore può avere rilevanza morale?**

Lo sviluppo di macchine senza guidatore pone importanti questioni di rilevanza morale. Innanzi tutto, è legittimo domandarsi quali doveri abbiamo verso queste macchine. Noi abbiamo una responsabilità anche nei confronti delle macchine non perché con i nostri comportamenti possiamo promuovere o compromettere il loro benessere, ma perché il loro cattivo funzionamento può danneggiarci, anche seriamente. Che l'autobus oppure l'autoveicolo vadano a fuoco non ci interessa più di tanto - a prescindere dal costo, in fondo si tratta di macchine e possiamo sempre sostituirle -, ma l'esplosione di una autovettura può ferire oppure uccidere persone innocenti. Per questa ragione, siamo tenuti a controllare le nostre macchine periodicamente (sia l'efficienza della performance che i livelli di emissioni inquinanti e rumorose), in quanto dobbiamo ridurre il rischio che provochino danni o incidenti evitabili. Ad ogni modo, le macchine in quanto tali hanno un valore semplicemente strumentale: mancano, cioè, di un valore intrinseco, indipendente dalla loro utilità. Tuttavia, è legittimo chiedersi se in futuro con lo sviluppo della robotica e dell'intelligenza artificiale

esse saranno ancora cose che possiamo usare come vogliamo. Negli ultimi decenni il nostro rapporto con gli animali è cambiato profondamente ed oggi sempre più persone ritengono sbagliato arrecare loro sofferenze. Quello che è successo negli ultimi anni con gli animali potrebbe succedere domani con robot intelligenti in grado di interagire con noi come se fossero autocoscienti? Noi possiamo sapere di avere davanti soltanto delle macchine tecnologicamente avanzate, programmate per interagire con noi, simulando il nostro comportamento, ma possiamo attribuire loro una vita mentale e le nostre stesse emozioni. Questo è un fenomeno che osserviamo già oggi con le tecnologie come ad esempio quelle impiegate in medicina, che non possiedono ancora elevati livelli di intelligenza: cosa accadrà quando con lo sviluppo scientifico e tecnologico ci troveremo ad interagire e a confrontarci con macchine sempre più intelligenti e (almeno apparentemente) capaci di provare le nostre stesse emozioni e sentimenti? Come potremo costruire relazioni con loro (e raccontare loro, ad esempio, i nostri pensieri e quello che proviamo), ma considerarle soltanto dei semplici oggetti [16]? Le macchine non sono persone, ma forse potremmo immaginare che lo siano.

Non stiamo ancora immaginando che un giorno le macchine diventeranno autocoscienti e che di conseguenza saranno in grado di provare i nostri stessi sentimenti. Quello che intendiamo affermare è che anche se le macchine che avremo davanti non proveranno 'veramente' i sentimenti e le emozioni che esprimeranno, questo potrebbe comunque non impedirci di considerarle l'altro di una relazione autentica [17]. È vero che una cosa sono i cosiddetti 'robot sociali' – cioè i robot autonomi o semi-autonomi in grado di interagire in maniera simile agli esseri umani, seguendo regole sociali o comportamenti legati al loro ruolo specifico – un'altra cosa le macchine che utilizziamo come mezzo di trasporto per muoverci da una località all'altra. Tuttavia, anche il software delle auto senza guidatore potrebbe essere in grado di conversare con l'utente, rispondere alle sue domande ed essere sensibile al suo stato d'animo. Se ipotizziamo, cioè, che anche le macchine senza guidatore saranno programmate per interagire come qualsiasi assistente, anche attraverso l'espressione di emozioni, per qualsiasi utente potrebbe essere facile empatizzare con il guidatore virtuale.

Avere, poi, una relazione con la macchina autonoma diventa ancora più facile, se immaginiamo che, a livello di software, la guida possa essere controllata dai dispositivi che impieghiamo quotidianamente a casa oppure a lavoro come assistenti vocali, in quanto avremmo già costruito una relazione con il nostro interlocutore virtuale. Qualcuno potrebbe essere scettico riguardo alla possibilità di attribuire una personalità ad un interlocutore che è presente nella nostra vita soltanto con la sua voce, ma questi dispositivi che oggi sono trasportabili, domani potrebbero essere integrati a casa oppure a lavoro con macchine in grado di muoversi ed interagire con noi come robot. Naturalmente, più tratteremo le macchine come persone e degne del nostro rispetto morale e più si apriranno dilemmi che al momento sembrano scenari di fantascienza. In caso di pericolo, ad esempio, chi dovremmo salvare: un umano o una macchina? Qualcuno potrebbe essere tentato di salvare la macchina con la quale ha costruito una relazione e che gli è stato vicino nei momenti più importanti della

propria vita. Per questa ragione, alcuni ritengono che sarebbe meglio che i robot non assomigliassero agli esseri umani perché i proprietari o, comunque, gli utenti dovrebbero aver sempre chiaro che essi non hanno alcun obbligo nei loro confronti [18].

### 3. Una macchina senza guidatore è un agente morale?

Inoltre, le macchine autonome ovverosia senza guidatore interessano la moralità perché la loro progettazione non può prescindere dalla progettazione di una 'competenza' etica. Queste macchine, infatti, non soltanto dovranno essere affidabili in termini di sicurezza e contribuire poi ad una riduzione degli incidenti su strada, ma dovranno anche essere in grado di risolvere i dilemmi morali che di volta in volta si presenteranno. Qui la questione che innanzi tutto si pone è se una macchina intelligente, progettata per comportarsi bene, possa essere considerata un 'agente' morale. Se assumiamo che la morale abbia a che fare con la ragione la risposta è scontata. Una macchina senza guidatore non può essere un agente morale perché comportarsi moralmente significa non soltanto distinguere tra comportamenti appropriati e non appropriati, ma anche essere capace di lasciarsi governare dalla ragione. Per questa ragione, afferma Kant, le regole della moralità valgono universalmente e si applicano a qualsiasi individuo in grado di considerare le cose razionalmente [19]: la ragione, infatti, prescrive le stesse regole o principi a qualsiasi individuo [20]. Forse un giorno anche le macchine più intelligenti mostreranno la capacità di agire secondo quella massima che può valere come massima di una legislazione universale, oppure, in un'altra formulazione, come legge 'universale' della natura. Tuttavia, fino a quando non saranno in grado di determinare da sole la loro volontà (ovverosia, non saranno capaci di darsi da sé la 'legge' o il principio della condotta) le loro scelte non saranno morali, in quanto saranno il risultato della nostra progettazione. Noi possiamo essere consapevoli della legge morale perché abbiamo una ragione che genera essa stessa le leggi o principi a cui volontariamente ci sottoponiamo: una macchina, invece, può anche essere in grado di imitare il nostro comportamento, ma manca della ragione e di conseguenza non può volere agire moralmente. Al massimo, una macchina potrebbe ingannarci con comportamenti, apparentemente morali, ma una macchina del genere non potrebbe essere un 'agente' morale. È vero che non possiamo escludere che domani riusciremo a produrre macchine così tecnologicamente avanzate da essere consapevoli di sé – cioè, autocoscienti [21]: almeno per il momento, però, si tratta di un'ipotesi remota e quasi fantascientifica.

Tuttavia, le cose cambiano se assumiamo che la ragione abbia un ruolo estremamente limitato nell'ambito della moralità e che le ragioni motivanti appartengano a sentimenti ed a passioni che sfuggono a qualsiasi controllo della ragione. Questa posizione è riassunta nel celebre passaggio del *Trattato sulla natura umana*, in cui Hume afferma che la ragione è – ed è giusto che sia – schiava delle passioni [22]. In questo caso, il fatto che la macchina manchi dell'autonomia della volontà, nel senso che non sia in grado di assoggettarsi ad una legislazione universale, non sarebbe più un motivo sufficiente per non



considerarla un agente morale, in quanto nessuno di noi ha una volontà che si lascia guidare dalla propria ragione. Per altro, è vero che l'agire della macchina sarebbe programmato da terze persone, ma anche il nostro comportamento è il risultato di processi educativi attraverso i quali impariamo ad apprezzare alcuni comportamenti ed a disprezzarne altri. Del resto, «(n)on è quindi una differenza da poco – afferma Aristotele –, se fin dalla nascita veniamo abituati in un modo piuttosto che in un altro, è importantissima, anzi è tutto» [23]. L'educazione che riceviamo, infatti, quando siamo ancora molto piccoli e non siamo ancora in grado di riflettere, crea abitudini e ci indirizza nella direzione giusta: «si deve essere abituati in un certo modo fin da giovani, come dice Platone, in modo da godere e rammaricarsi per le cose per cui è doveroso provare questi sentimenti...» [23]. In teoria, abbiamo sempre la possibilità di cambiare il nostro carattere, ma si tratta di una possibilità limitata perché «anche per l'ingiusto o l'intemperante vi era all'inizio la possibilità di non diventare persone di quel tipo – e per questo lo sono volontariamente – ma quando lo sono diventati non hanno più la possibilità di non esserlo» [23]. È vero, infine, che un agente morale può essere influenzato dal giudizio di approvazione o disapprovazione dei propri simili: anzi, si potrebbe affermare che la moralità ha proprio la funzione di incoraggiare e scoraggiare a compiere determinate azioni. Noi ci sforziamo ogni giorno, afferma Adam Smith, per fare in modo che il nostro comportamento sia in linea con i giudizi e i sentimenti di quello spettatore imparziale che vive dentro di noi ma che non rappresenta altro che la nostra natura sociale. Tuttavia, anche una macchina potrebbe essere sensibile ai giudizi delle altre persone: dovremmo solamente programmarla ad impostare il suo comportamento sulla base delle risposte (approvazione e/o disapprovazione) che riceve. Forse una macchina non sarà mai in grado di vergognarsi o essere orgogliosa per quello che ha fatto, ma potrebbe agire come se provasse questi sentimenti.

#### **4. Come programmiamo una macchina ad agire moralmente?**

Resta aperta, comunque, la questione di come programmare moralmente una macchina: cioè, quale concezione della moralità e quali valori dovremmo prendere a modello? In qualsiasi momento, infatti, una macchina potrebbe trovarsi di fronte a scelte importanti che possono riguardare altre macchine ma anche gli esseri umani [24,25,26]. Immaginiamo che la macchina senza guidatore debba scegliere se “sterzare a sinistra e colpire una bambina di otto anni, o sterzare a destra contro una nonna di ottanta anni. Data la velocità della macchina, la vittima sarebbe sicuramente uccisa dall'impatto. Se tu non sterzi, entrambe le vittime saranno colpite e nello scontro moriranno: così abbiamo buone ragioni di pensare che tu devi sterzare in un modo o nell'altro. «Ma – si chiede Patrick Lin – qual è la decisione eticamente corretta? Se tu stesso avessi la responsabilità di programmare la macchina, che cosa le faresti fare, se mai dovesse capitarle di incontrare un caso simile, non importa quanto raro possa essere?» [27]. A prima vista, la cosa più corretta da fare sembrerebbe quella di salvare la bambina, che ha davanti a sé ancora una lunga vita da vivere – ad

esempio, un primo amore, la costruzione di una sua propria famiglia, una carriera e la sua giusta quota di esperienze. Tuttavia, perché la nonna non dovrebbe aver diritto alla stessa rilevanza morale? Se, cioè, è sbagliato discriminare le persone per il genere, la religione o la disabilità, perché dovremmo avere il diritto di trattare le persone diversamente a seconda dell'età? Nemmeno l'altra possibile soluzione, però, sembra eticamente accettabile: potremmo programmare la macchina per non sterzare né a destra né a sinistra, in questo modo, però, condanniamo ad una morte certa sia la bambina che la nonna. Si pensi, poi, a quest'altro scenario: una macchina procede lungo la strada con il proprietario ed altri passeggeri a bordo, proprio al passaggio della macchina, improvvisamente un bambino attraversa la carreggiata per andare a riprendere il pallone. In una situazione di questo tipo cosa dovrebbe fare la macchina senza guidatore? Dovrebbe proseguire per la propria strada ed eventualmente investire anche il bambino, o dovrebbe cercare di evitare la morte del bambino con una manovra sull'altra corsia, anche al rischio in questo modo di causare un incidente con altre macchine ed eventualmente la morte o il grave ferimento di un numero maggiore di persone? È naturale pensare che nel caso in cui la manovra non comportasse alcun altro incidente, sarebbe giusto che la macchina cercasse di evitare di uccidere il bambino: il problema è che sterzando la macchina potrebbe causare un incidente molto più drammatico. Ad esempio, a seguito della manovra, la macchina potrebbe finire contro un autobus: l'autobus è pesante, ma lo scontro con l'automobile potrebbe farlo uscire di strada, si potrebbe ribaltare e poi prendere fuoco, ed i passeggeri non avrebbero scampo. A fronte di questi importanti dilemmi morali, possiamo scegliere di programmare la macchina a seguire un'etica deontologica oppure una morale consequenzialista [28]. L'etica deontologica afferma che un'azione o una scelta è giusta se è conforme ad una regola morale ed è sbagliata se, invece, è in contrasto con una norma morale: queste norme, inoltre, sono valide a prescindere dal fatto che promuovano il bene (ad esempio, la felicità delle persone coinvolte, il soddisfacimento delle loro preferenze oppure qualsiasi altra cosa a cui noi possiamo riconoscere una qualche valore), sono basilari e non sono derivate per induzione da giudizi su casi particolari. Di fatto – afferma Frankena – i sostenitori di un'etica deontologica asseriscono che dobbiamo sempre decidere che cosa fare alla luce di queste norme o principi [29]. Un'etica consequenzialista, invece, sostiene che la giustezza o scorrettezza di un'azione o di una scelta possa essere stabilita soltanto sulla base delle sue conseguenze. Una particolare versione di un'etica consequenzialista è, poi, l'utilitarismo che afferma che il giusto consiste nella promozione del bene generale: «(q)uesto principio afferma decisamente che il fine morale da ricercare in tutto quanto facciamo è la maggiore rimanenza possibile del bene sul male (o la minor rimanenza possibile di male sul bene) nel mondo nel suo complesso. (...) Questo implica che in ogni caso, qualunque cosa siano, il bene e il male sono tali da essere misurabili e confrontabili...» [29].

Se scegliamo un'etica utilitarista dovremmo programmare la macchina per ridurre – in caso di incidente o di scelta morale – il numero di persone ferite e uccise. Oppure, nel caso fosse in gioco lo stesso numero di persone, per salvare la persona più giovane (che ha cioè più anni da vivere) o con una



qualità della vita migliore. Un'etica deontologica, invece, ci vieterebbe sempre di 'sacrificare' persone innocenti, a prescindere dai vantaggi che si potrebbero ottenere in termini di vite salvate, in quanto, in questo caso, la cosa doverosa è rispettare la norma di non uccidere. Il principale problema di avere in circolazione macchine programmate con un'etica deontologica è che esse potrebbero interpretare le norme in maniera troppo letterale. Ad esempio, ipotizziamo che sia possibile programmare le macchine senza guidatore con le tre leggi della robotica di Asimov: 1) un robot non può arrecare un danno ad un essere umano, oppure attraverso il mancato intervento, lasciare che un essere umano sia danneggiato; 2) un robot deve obbedire agli ordini degli esseri umani tranne quando questi ordini sono in conflitto con la prima legge; 3) un robot deve proteggere la sua propria esistenza fintantoché ciò non contrasta con la prima e seconda legge. Davanti al pericolo di una collisione, la macchina potrebbe dare un'interpretazione troppo letterale di queste tre leggi e non frenare per evitare traumi ai passeggeri. Ma per le stesse ragioni la macchina potrebbe rifiutarsi di entrare nel traffico cittadino o di raggiungere certe destinazioni, in quanto 'sentirebbe' la responsabilità di minimizzare il rischio di incidente per il conducente e/o le altre persone. Inoltre, più aumentano le regole e i principi morali che bisogna rispettare, più aumenta la probabilità che, in particolari situazioni, due o più principi entrino in conflitto[30]: dovremmo poi programmare le macchine autonome per affrontare queste situazioni e per "sapere" qual è, in caso di conflitto, il principio che deve prevalere. A mano a mano che si presentano nuovi scenari e nuove circostanze dovremmo pertanto aggiungere nuove regole e principi e specificare la loro applicazione. Anche questo, però, non sarebbe risolutivo, in quanto, anche ammesso che sia possibile specificare sempre quando una norma ha la precedenza sull'altra, con lo sviluppo tecnologico, le auto potrebbero trovarsi in situazioni che non avevamo previsto. «Anche se l'etica deontologica può fornire una guida in molte situazioni, non è impiegabile come sistema etico completo a causa dell'incompletezza di ogni insieme di regole e della difficoltà di articolare l'etica umana nella sua complessità in una lista di regole» [31].

Sembra, pertanto, preferibile affidarsi ad un'etica che abbia un unico principio morale. Tuttavia, anche un'etica consequenzialista presenta una serie di problemi: immagiamo, ad esempio, di voler programmare le macchine secondo un'etica utilitaristica. Noi vogliamo che le macchine autonome riducano il più possibile il danno: ma come si calcola il danno? Soltanto in termini di vite umane coinvolte? Oppure consideriamo anche il danno economico (quale sarà il costo complessivo dell'incidente)? Inoltre, per calcolare il danno dobbiamo tener conto soltanto delle conseguenze più immediate o prendere in considerazione anche gli effetti a lungo termine. Nel primo caso, potendo scegliere se scontrarsi con un motociclista con il casco o un motociclista senza casco, la macchina dovrebbe optare per quello con il casco, in quanto il motociclista con il casco avrebbe maggiori probabilità di sopravvivere. In questo modo, però, si scoraggerebbe l'uso del casco e/o si penalizzerebbero le persone che rispettano la legge ed adottano una guida molto più prudente. Per le stesse ragioni, poi, la macchina autonoma dovrebbe preferire un incidente con una macchina più robusta (ad esempio, un SUV) rispetto ad una più leggera, in

quanto in caso di incidente la macchina più robusta sarebbe molto più resistente e perciò il rischio che i passeggeri siano feriti ed uccisi sarebbe molto più basso. A quel punto, però, chi avrebbe ancora la voglia di comprarsi questa macchina? Inoltre, si tratterebbe di decidere che prospettiva la macchina autonoma dovrebbe assumere: se proteggere sempre gli interessi del proprietario o "conducente": oppure assumere un punto di vista più fermo e generale e proteggere gli interessi della collettività? Immaginiamo ad esempio quest'altra situazione: la macchina potrebbe procedere nel modo più sicuro per il suo proprietario se, ad esempio, investisse il bambino che attraversa improvvisamente la strada per andare a riprendere il suo pallone. In fondo per qualsiasi macchina è meglio scontrarsi con l'"oggetto" più leggero: ad esempio, per lei sarebbe meglio scontrarsi con una moto piuttosto che un SUV. Se pertanto la scelta è tra un bambino e una macchina o addirittura un camion, la macchina avrebbe maggiori possibilità di uscirne indenne se colpisse ed investisse il bambino. L'alternativa è assumere un'altra prospettiva e sterzare su un'altra corsia, perché in questo modo si potrebbe salvare la vita del bambino senza conseguenze collaterali. C'è il rischio, però, che la manovra causi la morte del proprietario della macchina: sterzando, infatti, la macchina potrebbe uscire fuori strada e ribaltarsi, o oltre la strada c'è una foresta e c'è la possibilità che la macchina finisca contro un albero. Se consideriamo le cose moralmente, sembra giusto che la macchina non si preoccupi soltanto del suo proprietario o conducente, ma assuma un punto di vista generale: contano i miei interessi ed il mio benessere, ma anche quelli di qualsiasi persona. Tuttavia, anche in questo caso certe decisioni potrebbero rivelarsi controproducenti: ad esempio, programmare le macchine autonome per proteggere l'interesse generale potrebbe ridurre l'interesse per queste macchine e rallentare la loro diffusione. Le persone, infatti, potrebbero non sentirsi più al sicuro nelle loro autovetture o potrebbero preferire viaggiare su macchine su cui hanno un maggiore controllo. Anche questi sono scenari che un'etica consequenzialista dovrebbe considerare.

Inoltre, una macchina programma a scegliere sempre secondo un'etica utilitaristica e, di conseguenza, per promuovere la massima felicità per il più grande numero di persone potrebbe optare per soluzioni che possono apparire profondamente ingiuste. Pertanto, il criterio per determinare il giusto non può essere pertanto soltanto l'utilità: «(se) la giustizia può avere di tanto in tanto il sopravvento sull'utilità, allora – come ricorda Frankena – la questione di cosa sia giusto non può trovare risposta appellandosi al principio di utilità e i deontologi hanno, dopo tutto, ragione almeno in parte» [29]. In altri termini, entrambe le concezioni etiche presentano vantaggi e svantaggi: «Gli approcci deontologici richiedono regole astratte che possono non essere applicabili o che possono trovarsi in conflitto in particolari situazioni, mentre gli approcci consequenzialisti sono così rigidi che producono azioni che appaiono criticabili» [31]. Tuttavia, la programmazione di una macchina secondo una etica deontologica sembra impraticabile soprattutto per difficoltà interne a questa concezione: incompletezza di ogni sistema di regole e di qualsiasi graduatoria che intenda specificare, nei diversi casi, quale norma o principio morale debba avere la priorità. Anche il ricorso ad una concezione etica consequenzialista presenta

problemi, ma le difficoltà sembrano prevalentemente pratiche e di conseguenza superabili attraverso una selezione condivisa degli obiettivi che la macchina autonoma deve perseguire. Superare il disaccordo può richiedere tempo perché in una società pluralistica, come la nostra, abbiamo concezioni del bene e della giustizia molto diverse. Ma questo risultato può essere raggiunto con il coinvolgimento delle diverse parti in causa (popolazione, ingegneri, programmatori, industrie automobilistiche, ecc.) ed un confronto sui valori che vogliamo vedere rappresentati a livello di progettazione [32, 33, 34, 35].

## Conclusioni

In molti ambiti della vita e delle professioni l'inarrestabile avanzata di robot sempre più intelligenti a spese degli umani solleva importanti questioni morali e suscita preoccupazioni importanti riguardo alla capacità delle macchine di rimpiazzare gli umani. Ad esempio, c'è il timore che una cura amministrata da robot non sarebbe più umana in quanto impedirebbe al disabile e al paziente di coltivare importanti spazi di socialità. C'è la preoccupazione, poi, che l'impiego nei conflitti armati di macchine sempre più autonome, possa portare ad una escalation senza precedenti della violenza, in quanto la guerra non sarebbe più combattuta dagli esseri umani, ma semplicemente dai robot. Con l'impiego dei robot, inoltre, la guerra potrebbe perdere la sua umanità, in quanto una macchina intelligente non può essere mossa dalla compassione o dall'empatia. Nel caso, poi, dei sex robot è stato affermato che il loro sviluppo incoraggerebbe la violenza sulle donne rappresentandole come corpi sempre a disposizione dei maschi. Non sembra, invece, che perdiamo qualcosa di importante se smettiamo di guidare o deleghiamo, come finora abbiamo fatto, la guida non ad un conducente di professione (ad esempio, un autista dell'autobus o ad un taxista) ma ad un sistema operativo. Probabilmente, alcune persone si lamenteranno perché non potranno più guidare, ma i benefici a breve e lungo termine per la società potrebbero essere considerevoli, in quanto la guida autonoma potrebbe ridurre il numero di morti sulle strade. Ogni anno, del resto, 1,25 milioni di persone muore a causa di incidenti automobilistici e all'incirca il 90% di questi incidenti dipende da un errore umano [10,11].

Oggi diversi gruppi di ricerca lavorano per rendere i veicoli a guida autonoma sempre più sicuri, ad esempio, attraverso avanzati sistemi di controllo laterale e longitudinale, di intersezione, algoritmi di pianificazione del percorso con specifici requisiti [36]. Le autovetture, però, dovrebbero essere in grado non soltanto di identificare il pericolo (ad esempio, un'altra macchina che sta arrivando sulla corsia opposta, oppure un oggetto che si trova lungo la carreggiata) ma anche reagire prontamente. Un risultato che si cerca di raggiungere predisponendo la macchina a cambiare velocità e, se possibile, anche frenare davanti all'ostacolo in tempi brevissimi [37]. Affinché, però, possano risultare veramente affidabili per i passeggeri e terzi, le macchine autonome dovrebbero essere anche in grado di comportarsi moralmente: questa è la sfida più impegnativa per chi lavora allo sviluppo di questi autoveicoli. Naturalmente se con i veicoli autonomi si potessero eliminare tutti gli incidenti, i problemi sarebbero risolti perché non ci sarebbero più scelte morali

da fare: questo, tuttavia, non sembra – almeno per il momento – lo scenario futuro più realistico. Pertanto, non abbiamo altra scelta: dobbiamo rendere le macchine morali, almeno quanto basta per fare scelte che possano essere considerate accettabili dalla società. Fino ad allora, le macchine senza guidatore saranno semplicemente una speranza.

## Bibliografia

- [1] Hashim H. H., Omar, M. Z. (2017). "Towards autonomous vehicle implementation: issues and opportunities", *Journal of the Society of Automotive Engineers Malaysia*, 1, 2, 111-123.
- [2] Karnouskos S. (2018). "Self-driving car acceptance and the role of ethics", *IEEE Transactions on Engineering Management*, 1-15.
- [3] Burns L. D. (2013). "Sustainable mobility: a vision of our transport future", *Nature*, 497, 181-182.
- [4] Fagnant, D. J., Kockelman K. M. (2014). Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations for capitalizing on self-driven vehicles, paper presented at the 93rd Annual Meeting of the Transportation Research Board, Washington DC.
- [5] Fagnant, D. J., Kockelman, K. M. (2014). "The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios", *Transportation Research Part C: Emerging Technologies*, 40, 1-13.
- [6] Sivak, M., Schoettle, B. (2015). *Motion sickness in self-driving vehicles*, University of Michigan Transportation Research Institute.
- [7] Litman T. (2019). "Autonomous vehicle implementation predictions implications for transport planning", Victoria Transport Policy Institute, <https://www.vtpi.org/avip.pdf> (ultimo accesso 25 settembre 2019).
- [8] Litman T. (2014). "Ready or waiting", *Traffic Technology International* ([www.trafficechnologytoday.com](http://www.trafficechnologytoday.com)), 36-42; [http://www.vtpi.org/AVIP\\_TTI\\_Jan2014.pdf](http://www.vtpi.org/AVIP_TTI_Jan2014.pdf) (ultimo accesso 25 settembre 2019).
- [9] Sparrow, R., Howard M. (2017). "When human beings are like drunk robots: driverless vehicles, ethics, and the future of transport", *Transportation Research Part C: Emerging Technologies*, 80, 206-215.
- [10] World Health Organization (2015), *Global status report on road safety 2015*, Switzerland, World Health Organization.
- [11] Anderson, J. M., Nidhi, K., Stanley, K. D., Sorensen, P., Samaras, C., Oluwatola, O. A. (2016). *Autonomous vehicle technology: A guide for policymakers*, Rand Corporation, Santa Monica.
- [12] Pakusch C. et al. (2018). "Unintended effects of autonomous driving: a study on mobility preferences in the future", *Sustainability*, 10, 2404, 1-22.
- [13] Ferreras, L. E. (2013), "Autonomous vehicles: a critical tool to solve the XXI century urban transportation grand challenge", in Jones, S. L. (a cura di), *Urban public transportation system 2013: proceedings of the third international*

conference on urban public transportation systems, 17-20 November 2013, Paris, 405-412.

[14] Levin, M. W., Boyles, S. D., Patel, R. (2016). "Paradoxes of reservation-based intersection controls in traffic networks", *Transportation Research Part A: Policy and Practice*, 90, 14-25.

[15] Levin, M. W., Boyles, S. D. (2015). "Intersection auctions and reservation-based control in dynamic traffic assignment", *Transportation Research Record: Journal of the Transportation Research Board*, 2497, 35-44.

[16] Dumouchel, P., Damiano L. (2019). *Vivere con i robot. Saggio sull'empatia artificiale*, Raffaello Cortina, Milano.

[17] Turkle, S. (2012). *Insieme ma soli. Perché ci aspettiamo sempre più dalla tecnologia e sempre meno dagli altri*, Codice, Torino.

[18] Bryson J. J. (2010). "Robots Should Be Slaves", in Wilks Y. (a cura di), *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issue*, John Benjamins, Amsterdam, 63-74.

[19] Kant I. (1989). *Critica della ragion pratica*, Laterza, Roma-Bari.

[20] Driver J. (2007). *Ethics. The Fundamentals*, Blackwell Publishing, Oxford, 80-101.

[21] Tegmark M. (2018). *Vita 3.0. Esseri umani nell'era dell'intelligenza artificiale*, Raffaello Cortina, Milano.

[22] Hume D. (1982). *Trattato sulla natura umana*, Laterza, Roma-Bari.

[23] Aristotele (1999). *Etica Nicomachea*, Laterza, Roma-Bari 1999.

[24] Nihols S., Smids J. (2016). "The ethics of accident-algorithms for self-driving cars: an applied trolley problem?", *Ethical Theory and Moral Practice*, 19, 5, 1275-1289.

[25] Renda A. (2018). "Ethics, algorithms and self-driving cars – a CSI of the 'trolley problem'", *CEPS*, 2, 1-17.

[26] Sabin D. (2017). "New algorithm lets self-driving cars merge with traffic like a school of fish", <https://www.inverse.com/article/27119-algorithm-merge-autonomous-highway>.

[27] Lin P. (2015). "Why ethics matters for autonomous cars", in Maurer, M., Gerdes, J. C., Lenz, B., Winner, H. (a cura di) *Autonomous driving: technical, legal and social aspects*, Springer, Berlin, 69-85.

[28] Bonnefon, J. F., Shariff, A., Rahwan, I. (2016). "The social dilemma of autonomous vehicles", *Science*, 352, 1573-1576.

[29] Frankena W. K. (1996), *Etica. Un'introduzione alla filosofia morale*, Edizioni di Comunità, Milano.

[30] Li S., Zhang, J., Wang, S., Li, P., Liao, Y. (2018). "Ethical and legal dilemma of autonomous vehicles: study on driving decision-making model under the emergency situations of red-light running behaviors", *Electronics*, 7, 264, 1-18.

- [31] Goodall N. (2014). "Ethical decision making during automated vehicle crashes", *Transportation Research Record: Journal of the Transportation Research Board*, 2424, pp. 58-65.
- [32] Etzioni A., Etzioni O. (2017). "Incorporating ethics into artificial intelligence", *Journal of Ethics*, 21, 4, 403-418.
- [33] Holstein T. (2017). "The misconception of ethical dilemmas in self-driving cars", *Proceedings*, 1, 174, 1-3.
- [34] Goodall, N. J. (2016). "Can you program ethics into a self-driving car?", *IEEE Spectr.*, 53, 28-58.
- [35] McBride, N. (2016). "The ethics of driverless cars", *SIGCAS Comput. Soc.*, 45, 179-184.
- [36] Milakis, D., van Arem, B., Van Wee, B. (2017), "Policy and society related implications of automated driving: a review of literature and directions for future research", *Journal of Intelligent Transportation Systems*, 21, 4, 324-348.
- [37] Bichiou, Y., Rakha, H. A. (2019), "Real-time optimal intersection control system for automated/cooperative vehicles", *International Journal of Transportation Science and Technology*, 8, 1, 1-12.

## Biografia

**Maurizio Balistreri** è Ricercatore di filosofia morale presso il Dipartimento di Filosofia e Scienze dell'Educazione dell'Università di Torino. Ha collaborato al *Dizionario di bioetica* curato da Eugenio Lecaldano (2000), è autore di numerosi saggi di bioetica e di *Etica e romanzi* (Le Lettere, 2010), *Superumani: etica ed enhancement* (Espress edizioni 2011), *La clonazione umana prima di Dolly* (Mimesis 2015) *Il futuro della riproduzione umana* (Fandango 2016) e *Sex robot. L'amore al tempo delle macchine* (Fandango 2018; Rowman & Littlefield International 2020).

Twitter: @m\_balis

Email: maurizio.balistreri@unito.it